Transformative Decisions

Kevin Reuter & Michael Messerli

forthcoming in the *Journal of Philosophy*, pre-review version

Some of the most fundamental decisions we make in our lives – like becoming a parent or moving to a different part of the world – are transformative. According to Laurie Paul (2014), transformative decisions pose a major problem to us because they fall outside the realm of rationality. Her argument in favor of that conclusion rests on the premise that the subjective value, i.e. the value of experiencing a certain outcome of a decision, plays the central role in transformative decisions. This paper challenges that premise and hence the overall conclusion that transformative decisions are usually not rational. In the theoretical part, we specify the conditions under which transformative decisions are *possibly* rational and *likely* to be rational. The data we present in the empirical part reveals that the subjective value often plays only a minor role in people's decision process. Putting both parts together, we argue that people have a great chance in making rational transformative choices.

Keywords: Transformative Experience; Subjective Value; Rational Choice; Multi-criteria Decision Making; Experimental Studies.

1 Introduction

Imagine that you are on holidays and you are confronted with a choice between two Indian dishes that are made of spices you have never tasted before. Suppose you consider to have a child or remain childless. Or, slightly less mundane, think about aliens that have taken over the world, and you have to make a choice between either being transported to an unknown planet

where you can start a new life, or staying on earth but living under unknown alien rule.

In her book *Transformative Experience*, Paul argues that we usually cannot make a rational choice in such decision making situations (Paul 2014). The first part of her argument rests on two premises. Premise (i) concerns the nature of the experiences that characterize those decision – hereafter transformative decisions. Transformative decisions involve either epistemically (spices case), or both epistemically and personally transformative experiences (child case, alien rule case), where having an epistemically transformative experience of x is *the only way* to know what it is like to x, e.g., having the experience of being a father or eating dragon fruit is the only way to know what it is like to be a father or to eat dragon fruit. A transformative experience which is also *personally* transformative, additionally changes one's fundamental preferences, e.g., experiencing life on an alien planet will certainly change one's core preferences.[1] Importantly, when a decision is epistemically transformative, the agent faces an epistemic wall which she cannot penetrate through either imagination, testimony or scientific investigations.[2]

Premise (ii) of the argument states that the *subjective value* – the value of experiencing a certain outcome of a decision – plays the central role in transformative decisions. Applying this point to the decision on whether or not to have a child, Paul claims: "(...) the value of what it is like for the agent plays the central role, if not the only role, in the decision to procreate."(Paul 2015a, p. 153). While this aspect of the subjective value is often neglected in discussions on transformative choices, it is a crucial ingredient to Paul's argument and is emphasized by Paul frequently.[3] Consider an example of Paul

[1] Purely personally transformative experiences are not discussed at any length by Paul.

[2] Paul states: "(...) and remember, current science doesn't pronounce on how you'll respond, and experiences vary widely, so you can't rely on testimony."(Paul 2014, p. 36). Paul's main point is that scientific data on large groups do not give fine-grained information on how *you* respond to a particular experience (Paul 2014, p. 131, 132). See Cappelen & Dever (2016), for a critical discussion how empathy is related to the ideology of the first person.

[3] E.g. she writes: "(...) and an essential part of your deliberation concerns what your future

where the subjective value fails to play the central role: Many adults decide to have a child because they "need an heir, or more hands to work the farm, or whatever."(Paul 2014, p. 85) While this is perhaps not the most common approach for most adults in our times, the general point should be clear. Paul's challenge to the rationality of a transformative decision does not even begin to arise, if people make decisions independently of the transformative experiences connected with those decisions. Thus, the challenge only applies if the subjective value plays the central role. What this means exactly, will be clarified in section 2.

Given the first two premises of the argument, we can now substantiate Paul's claim that transformative choices cannot be rational, by considering the argument in full:[4]

i. One cannot know the subjective value of at least one outcome of a transformative decision d_t.[5]

ii. The subjective value plays the central role in transformative decisions d_t.

iii. If (i) & (ii), then the ranking of options in a transformative decision d_t cannot be determined.

iv. If the ranking of options in a decision cannot be determined, then such a decision cannot be rational.

$\therefore$ Transformative decisions d_t cannot be rational.

---

life will be like (...)"(Paul 2014, p. 3). For further discussions of this point, see also pages 4, 25, 26, 74, 75, 85, 126 of her book.

[4] See Friedman (unpublished manuscript), for an alternative formulation of the decision problem which focuses on the role of authenticity within transformative choices. See also Campbell (2015). However, it seems possible to formulate the challenge for rational choice without referring to the concept of authenticity.

[5] A transformative decision d_t is a decision which is not merely based on a non-experiential criterion. We use this specification to exclude those cases from the discussion in which the subjective value is fully disregarded in the decision making process.

One of Paul's most significant insights in her book is revealed by premise (iii) of the argument. If both premise (i) and premise (ii) are true, then it is impossible to determine an ordinal ranking of the possible options in such a decision.[6] If the subjective value of having a child cannot be identified, and if this value plays a central role in the decision process, then there is no way to assign a specific or even probable overall value to the decision option *having a child*. It follows that if no overall value can be assigned to that option, then ranking the options *having a child* and *not having a child* is rendered impossible. The final premise (iv) now states a basic tenet of standard rational decision theory. A decision can be rational only if the options in a decision process can be ranked.[7] The conclusion is as important as it is devastating. According to her argument, some of the most fundamental decisions in our life remain outside the realm of rationality. However, all hope is not lost. Paul suggests that instead of basing one's choice on the subjective value, we should instead make transformative decisions by asking ourselves how much we value new experiences or new selves (Paul 2014, p. 123).[8]

In this paper we challenge the soundness of Paul's argument by focussing primarily on premises (ii) and (iii), and present empirical results that suggest that premise (ii) is false for many transformative decisions. In contrast, we accept Paul's arguments in favor of premises (i) and (iv). In favor of premise (i), Paul draws an analogy with Jackson's famous Mary thought experiment: Not having seen colors so far in her life, Mary cannot know how it will feel

[6] Whereas Paul does not herself speak of ordinal ranking, we believe premise (iii) preserves the core idea of her claim. A similar reconstruction of that premise can be found in Dougherty et al. (2015, p. 305): "(...) If an agent cannot rationally judge the subjective value of a phenomenal outcome for her, then she cannot rationally choose between options when one of these options would lead to this phenomenal outcome."

[7] For a discussion of ordinal preference orderings, see, e.g. Savage (1954), Chapter 2. Also note, that a decision that cannot be rational is not thereby irrational, but rather falls outside the realm of rationality.

[8] See Shupe (2016) for a critical discussion of Paul's positive thesis. According to our empirical results, only in the vampire case do people primarily base their decisions on the revelatory value. Of course, the results do not tell us whether people *should* rely on the revelatory value (see section 3 and 4).

like to perceive colors until she leaves her room. Analogously, not having become a parent in one's life so far, we cannot know what it will feel like to become a parent. As noted above, premise (iv) is part of the standard normative decision theory, according to which agents should make a choice by determining the values of each possible outcome, and choose the outcome with the highest expected value. However, others have applied different concepts of rationality. First, according to Pettigrew, Paul does not challenge the *preference-first conception* (Pettigrew 2015, p. 668) which *merely* requires an agent to adhere to the basic axioms of decision theory, e.g. completeness, no transitivity violations.[9] Second, many theorists endorse a *satisfying theory of rationality*. Simply put, the idea behind this concept is that it is rational to choose an option in case it is sufficiently satisfactory.[10] Of course, one might argue that in a transformative choice an agent cannot fulfill the satisfying theory of rationality either, since she cannot know whether any of the options is good enough.

Having reconstructed Paul's argument and motivated its premises, here is how we will proceed in the rest of the paper: In Section 2, we analyze how the centrality of a decision criterion and the possibility to rank options in transformative decisions hang together. More specifically, we discuss what it means for the subjective value to be central such that the person will make a choice that is *necessarily* not rational. However, even if the subjective value of, e.g., becoming a parent, is not the central criterion, many transformative decisions will nonetheless fail to be rational. We therefore developed a statistical model that can tell us more precisely, how likely it is that a

[9] Does Pettigrew's challenge hold? One might at least doubt whether transformative decisions are *complete*: An agent might neither prefer to become a vampire nor to stay human nor to be indifferent between the alternatives. In this respect, there might exist a very interesting overlap with so-called hard cases of comparison (see, e.g., Chang (2015), Messerli & Reuter (2017).

[10] The concept goes back to Simon (1953). See also Weber (2000). In Simon's view, a notion of rationality according to which an agent needs to maximize utility is too demanding. That notion presupposes computational capacities that we do not possess.

decision is rational given the importance of the subjective value. Crucially, the likelihood increases with a decrease in the importance of the subjective value. Thus, to properly determine how likely it is that a person will make a rational transformative decision, we need to find out how people judge the importance of various decision criteria in transformative decisions. And that is exactly the task we have taken up in Section 3. In a series of studies, we not only found that premise (ii) of the argument seems mistaken, the data also suggests, that what it is like to experience a certain outcome is only one among many other decision criteria that play a role in transformative decisions. Section 4 concludes with a general discussion, whereby we deal with possible objections against our methodology and our results.

Before we proceed, we would like to address in advance one worry that readers familiar with Paul's writings might have. The experiments we have conducted reveal which decision criteria various people *in fact* take into account when facing a transformative decision. However, Paul's project is at least in part *normative*, i.e., she is concerned with normative decision theory, not with descriptive decision theory. Are we therefore not mistaken in challenging Paul's claim empirically? We are not, because the normative aspects of Paul's argument concern the relation between the values of the relevant outcomes and the rationality of the decision (premise (iv)) but not which criteria are considered important for determining the value of the outcomes. Thus, we fully agree with Paul when she states that "the normative standard for rational decision-making is that the agent or decision-maker *should* choose the act that has the highest expected value. (2014, p. 21, our italics). Indeed, in Section 2 we fundamentally rely on this normative standard to develop a model for determining the relation between the centrality of a decision criterion and whether or not the decision is rational. However, Paul makes a descriptive claim about the criteria that determine which outcome has the highest value. According to Paul, "while the subjective values are not the only values of these outcomes, they are some of the most central and important

ones, and an emphasis on them fits the dominant cultural paradigm."(2014, p. 26). Thus, Paul's claim and the putatively dominant cultural paradigm are not about how we should determine the value of certain outcomes, they are rather claiming that subjective values are most central and important. There can therefore be little doubt that Paul bases her normative claims on a purely descriptive analysis of the importance of subjective values.

2 Transformative Decisions and Rational Choice

In Section 1, we started our discussion of Paul's both surprising and insightful conclusion that transformative decisions d_t are not rational by reconstructing the underlying argument in favor of that conclusion. We also highlighted that a crucial premise in her argument is the claim that the subjective value(s) of the possible outcome(s) in a transformative choice is central to the decision. This claim immediately raises two questions: first, what does it mean to be central? Second, is it really the case that the subjective value plays the central role in transformative decisions. Paul does not address in very much detail the first question but primarily relies on our intuitive grasp of what it means to be central. In contrast, in regards to the second question, Paul contends that in a transformative choice like whether or not to become a parent, most people, at least in Western cultures, take the subjective value to be of utmost importance. In Section 3, we present results that strongly suggest that Paul is mistaken in her assessment of what most people take to be important in transformative decisions.

In this part of the paper, we will specify how we need to interpret "centrality" such that Paul's argument is valid (Section 2.1), and then provide a model that tells us how likely it is that a person makes a rational transformative choice given the importance that she places on the subjective value (Section 2.2).

2.1 When a transformative decision is possibly rational

To motivate our investigation into the topic, consider a hiring committee at a university that needs to decide which candidate to hire for an open academic position. Let us look at two extremes: On the one hand, it might be the case that the candidate's publication record is the only criterion that is taken into account. Subsequently, other factors like teaching skills, age, interests, etc. are of no concern. On the other hand, the committee might fully disregard the quantity and quality of the publication records, and, e.g., decides to select the person who lives closest to the university. Obviously, in the former case, a candidate's publication record is all that counts, and a fortiori central to the decision-making process. In the latter case, the publication record plays no role whatsoever. The reality of most decisions of hiring committees is hopefully in between these two extremes: Publication records are important but other criteria play a role too. Analogously, on the one hand, the what-it-is-like aspect of a certain outcome might be the only criterion in a transformative decision. In that case, the decision cannot be rational (assuming the truth of the other premises in Paul's argument). On the other hand, the what-it-is-like aspect might be of no concern in transformative decisions. Accordingly, the rationality of such transformative decisions is not threatened. Again, the reality of a transformative decision is likely to be somewhere in between these two extremes, i.e. most people when making a choice on whether to become a parent will consider various decision criteria like (a) financial situation, (b) what their partner wants, (c) whether having a child fits into one's vision for life, etc., and, perhaps also what it will feel like to have a child. Thus, the transformative decisions people face, will most likely be multi-criteria decisions. Paul acknowledges situations in which the subjective value plays no role at all, e.g., when discussing the king who decides to become a parent because he needs an heir to his throne. However, Paul fails to provide an answer to the question of how important the subjective

value must be when multiple criteria need to be evaluated.

An answer to this question can be easily given by making concrete what we have so far only considered in abstract terms. Thus, take Mary who contemplates the choice of becoming a parent. At first, Mary realizes that three criteria of varying importance are relevant to her. Most important to her is (a) whether her partner would like to have a baby. Less important but still of considerable weight is the question of (b) whether she and her partner can financially provide for a child. She also (c) imagines what it will be like for her to have a baby: while she thinks that the subjective value should be taken into account when making such a decision, the other two criteria play a greater role. We can put this in numerical terms by stating that the criteria priorities or *weights* w_i are as follows: $w_{partner} = 0.4$ (read this as saying that the criterion <partner> accounts for 40% of the decision process), $w_{costs} = 0.35$, $w_{whatitslike} = 0.25$.[11] The values for all criteria priorities should, of course, add up to 1, which is equivalent to saying that all decision criteria make up 100% of the decision process. In the second stage of the decision process, Mary determines the so-called local alternative priorities or *values* v_i, that is, how much the three criteria favor a certain outcome.[12] Let us first suppose that Mary's partner very much would like to have a child. Thus, for criterion <partner>, Mary assigns the value 1 to the outcome 'having a child', i.e. $v_{partner/pro\ child} = 1$, and 0 to 'not having a child' ($v_{partner/contra\ child} = 0$). Second, the financial situation of Mary and her partner is stable but not great. Thus, for criterion <costs>, Mary assigns the values $v_{costs/pro\ child} = 0.5$ and $v_{costs/contra\ child} = 0.5$. The subjective values $v_{whatitslike/pro\ child}$ and

[11] There are at least four common assumptions that most models of such decision processes make: (1) separability of the decision criteria (in order to avoid double-counting), (2) numerical ascertainability of the criteria weights, (3) weight and value attributions range between '0' and '1', and (4) for each i, $v_{outcome1}$+$v_{outcome2} = 1$. The empirical studies in the next section will demonstrate, how we arrived at the numerical weights of each decision-criterion. Furthermore, we will address objections against the separability of the decision criteria in the General Discussion.

[12] This is the usual terminology used in Multi-Criteria Decision Analysis (see e.g. Ishizaka & Nemery 2013, p. 16)

$v_{whatitslike/contra\ child}$ are unknown. By adding up the products of the weights and values, we can determine the overall utilities $U_{outcome} = \sum_{i=1}^{n} \mathrm{u}_i = \sum_{i=1}^{n} \mathrm{v}_i * w_i$ for each of the two outcomes given the first two decision criteria:

$$U_{pro\ child} = 1 * w_{partner} + 0.5 * w_{costs} = 1 * 0.4 + 0.5 * 0.35 = 0.575.$$

$$U_{contra\ child} = 0 * w_{partner} + 0.5 * w_{costs} = 0 * 0.4 + 0.5 * 0.35 = 0.175.$$

Now, even if the subjective value were to strongly speak against having a baby, e.g. $v_{whatitslike/contra\ child} = 1$, because Mary would feel completely drained by the additional responsibility and the lack of sleep, the overall utility of not having a child would be at most $U_{contra\ child} = 0.425$ $(0.175 + 0.25)$ and thus still much lower than $U_{pro\ child} = 0.575$. Thus, the what-it-is-like aspect of having a child has no bearing on the outcome of Mary's decision on whether or not to have a child. In other words, the subjective value is not important enough to influence the decision process: the transformative choice is rational. In order to guarantee that premise (iii) of Paul's argument is true, and, hence, that the transformative decision is not rational, the weight of the subjective value $w_{whatitslike}$ needs to be at least 0.5, i.e. it should at least account for 50% of the decision process. If the weight is equal or greater than 0.5, then even if all the other criteria speak for one outcome, the overall utility can flip in favor of the opposite outcome. If the weight is lower, however, then it is possible that the transformative decision is rational. We can therefore state more precisely what is meant by *central* in premise (ii) of Paul's argument: The subjective value plays the central role in a decision process if and only if it is at least as important as all other criteria taken together.[13]

How threatening are these considerations for the view that Paul holds?

[13]We do not claim that the way we have cashed out the notion of centrality matches Paul's concept of centrality. As mentioned above, Paul remains more or less silent on this matter.

There seem to be two possible ways to respond to the challenge we pose: First, one might maintain that the weight $w_{whatitslike}$ is generally greater than 0.5. The empirical data in Section 3 shows that such a response would be mistaken. Second, we might highlight that the value of 0.5 is only a cut-off point above which it is not possible to ordinally rank the outcomes. In almost all situations, however, a lower weight will still be sufficiently high. In the example above, $w_{whatitslike} = 0.25$ was not sufficiently high, but if the weight had been just a little higher, then no rational decision could have been made. In order to assess the strength of the second response, we now need to determine the dependency of the rationality of a transformative choice on the weights of the subjective value $w_{whatitslike}$.

2.2 When a transformative decision is likely rational

While in the last subsection, we focused on the *possibility* of making a rational transformative choice, in this section we will provide an answer to how likely it is that people make rational transformative choices. In the example above, we have seen that the criteria <partner> and <costs> were too important for the subjective value to play any role. Let us start by stating in more general terms when a transformative decision is rational. Assuming a binary choice, if one outcome is considered to be so much better than the other outcome, such that the subjective value will not make any difference, then the transformative decision is rational. It then simply does not matter whether the value $v_{whatitslike}$ can be determined or not. More precisely, we can state condition (R):

$$(R) \mid U_{outcome\ 1} - U_{outcome\ 2} \mid \geq w_{whatitslike},$$

where $U_{outcome} = \sum_{i=1}^{n} v_i * w_i$. (R) states that the absolute difference between the overall utility of one outcome and the overall utility of the other outcome –

without taking into account the subjective value – is greater than $w_{whatitslike}$. If (R) holds, then the transformative decision is rational.

Thus, the two types of variable that are required to determine whether or not a transformative decision is rational, are the weights of the criteria w_i as well as the values v_i. It seems we can now simply ask people to tell us the v_i and w_i of the various decision criteria and plug them into our formula. If Paul is correct in thinking that transformative decisions are in general not rational, we should find that most people violate condition (R), i.e. $w_{whatitslike}$ is sufficiently big to have a decisive influence on transformative decisions. In the next section, we present the results of several studies in which we asked hundreds of participants to report the weights w_i when being imaginatively confronted with various transformative decisions. However, we did not also determine the values v_i of the various decision criteria because this would have required of us to only consider people who currently face a transformative decision.[14] The online surveys we conducted allowed us to collect the w_i relatively easily from a random sample of people from the Amazon Mechanical Turk pool. While it would have been fascinating to interview people who are currently confronted with a transformative decision, this was not feasible for the purposes of this study.[15]

Are we therefore left without the means to connect the importance of the subjective value with the rationality of a transformative decision? No, fortunately, statistical models can be used to predict how likely it will be that a transformative decision is rational given the weight $w_{whatitslike}$. In our first statistical model, we simplified the complexity of the situation by assigning values v_i to each criterion by assuming that the values v_i can only take on values '1' and '0'. Thus, we assume that a decision criterion either

[14] To see why, consider again the choice of whether or not to have a child. Some decision criteria like <partner> are indeterminate for most participants unless they have a partner and are currently in the process of making such a decision.

[15] This restriction on our studies might fuel the worry that the participants' ratings are not representative of decisions that are made in actual decision-making contexts. We will respond to this worry in section 3 and section 4.

fully speaks in favor (1) of an option, e.g. having a child, or speaks against (0) the option, e.g. against having a child. Using the example from above, besides what it is like to feel being a parent, Mary considers two decision criteria to be relevant, <partner> and <costs>. Thus, there are $2^2 = 4$ possibilities in which the values can be combined: (1,1), (1,0), (0,1), (0,0). Using (R) above, we can therefore calculate how likely it is that John will make a rational decision.

1. $|1 * 0.4 + 1 * 0.35 - (0 * 0.4 + 0 * 0.35)| = 0.75 > 0.25$ (rational)
2. $|1 * 0.4 + 0 * 0.35 - (0 * 0.4 + 1 * 0.35)| = 0.05 < 0.25$ (not rational)
3. $|0 * 0.4 + 1 * 0.35 - (1 * 0.4 + 0 * 0.35)| = 0.05 < 0.25$ (not rational)
4. $|0 * 0.4 + 0 * 0.35 - (1 * 0.4 + 1 * 0.35)| = 0.75 > 0.25$ (rational)

Thus, in two out of four, i.e. 50% of the combinations, John will make a rational decision. In many transformative choices, there will be, of course, many more decision criteria that might be relevant to people's choices. In all the empirical studies we conducted (next section), we presented people with a total of six criteria including the subjective value. Thus, for the other five criteria, there are $2^5 = 32$ different possibilities that need to be calculated to determine the overall number of rational decisions giving the weights of all decision criteria, e.g. (1,1,1,1,1), (1,0,1,0,1).

We are now in a position to outline for some specific cases, the dependency of the rationality of a transformative decision on the subjective value $w_{whatitslike}$. Aside from using five decision criteria (excluding <what it is like>) and normalizing the values to 1, we first consider the case in which all weights (excluding the weight of the subjective value) are equal, i.e. $w_1 = w_2 = w_3 = w_4 = w_5 = (1 - w_{whatitslike})/5$. Starting with these assumptions, the number of rational transformative choices can be determined by the following formula:

$$\text{Number of Rational Decisions} = 2 * \ominus(1 - 2 * w_{whatitslike}) + 10 * \ominus(3 - 8 * w_{whatitslike}) + 20 * \ominus(1 - 6 * w_{whatitslike})$$

The plot of this function[16] for this very general case is also depicted in Figure 1 (upper-left hand side). Importantly, the graph shows that the less important the subjective value is (the smaller $w_{whatitslike}$), the greater the number of rational decisions, and hence the more likely that the person will indeed make a rational transformative decision. Given the assumptions we used, which simplified both weights w_i as well as values v_i, merely four different steps need to be distinguished. If $w_{whatitslike}$ falls below 0.5, there are two combinations for which the transformative decision is rational, namely the ones in which all other criteria speak in favor of the same outcome. However, if $w_{whatitslike}$ is lower than 0.375, then there are already 12 combinations for which the decision turns out to be rational. An even lower value of 0.16 or smaller ensures that the transformative decision is *necessarily* rational (all 32 possible combinations are rational decisions). These results might be interpreted to support Paul's view at least to some extent. After all, even if the subjective value does not play the central role, it might be important enough such that only very few combinations turn out to be rational. We will see that this *optimism* is premature.

For this admittedly simple model sketched above, we find that the lower $w_{whatitslike}$, the greater the number of rational decisions. But does this relation also hold, once we adopt a more realistic model? It does. To simulate a more complex situation, we dropped the assumption that all weights of

[16] $\ominus(x)$ is the Heaviside step function taking the value 0 if x < 0, and 1 otherwise. We give a brief explanation of how the first term of the formula can be derived. The other terms can be calculated in a similar fashion. In the case in which all non-experiential criteria speak in favor of one outcome, then the decision will be rational if $|w_1 + w_2 + w_3 + w_4 + w_5| \geq w_{whatitslike}$. We can now substitute $w_i = (1 - w_{whatitslike})/5$ and simplify the equation, which yields $1 - 2 * w_{whatitslike} > 0$. Thus, the condition is satisfied only if $w_{whatitslike}$ is smaller than 0.5. There are two possibilities that need to be considered: the case in which all non-experiential criteria speak in favor of one outcome, and the case in which they all speak against a certain outcome. Hence, the first term is $2 * \ominus(1 - 2 * w_{whatitslike})$.

the non-experiential decision criteria are assigned the same value. Rather, we stipulated weights of increasing value: $w_1 < w_2 < w_3 < w_4 < w_5$[17]. The upper-right hand side of Figure 1 shows how the number of rational decisions depends on the importance of $w_{whatitslike}$. Again, a decrease in the importance of the subjective value leads to an increase in the likelihood of making a rational transformative choice. As the lower part of Figure 1 shows, the general increase in rational decisions with decreasing $w_{whatitslike}$ neither depends on the normalization of the values (lower-left hand side; $v_1 = 0.3$, $v_2 = 0.7$, $v_3 = 0.3$, $v_4 = 0.7$, $v_5 = 0.3$), nor on having used five additional criteria (lower-right hand side, four criteria).

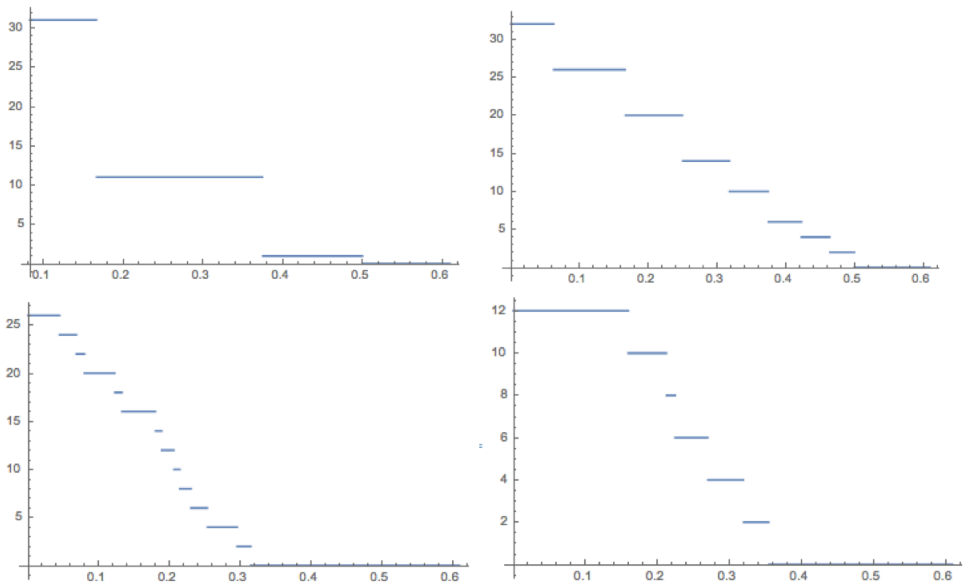

Figure 1: Plots of functions depicting the relation between the weight $w_{whatitslike}$ (X-Axis) and the number of rational decisions (Y-Axis). Upper left, weights w_i for all criteria except $w_{whatitslike}$ are equal. Values v_i are normalized; Upper rights, weights w_i for all criteria are different. Values v_i are normalized; Bottom left, weights w_i for all criteria are different. Values v_i are not normalized; Bottom right, weights w_i for all criteria are different. Only four additional criteria were used.

We have varied several parameters that play a role in determining the

[17]More precisely, we assigned the following priorities to each criteria $v_1 = 1$ * (1-$w_{whatitslike}$) / 15 < $v_2 = 2$ * (1-$w_{whatitslike}$) / 15 < $v_3 = 3$ * (1-$w_{whatitslike}$) / 15 < $v_4 = 4$ * (1-$w_{whatitslike}$) / 15 < $v_5 = 5$ * (1-$w_{whatitslike}$) / 15.

dependency of the number of rational transformative decisions on the subjective value. The results paint a clear picture: The likelihood that a person makes a rational transformative decision is greater, the smaller the weight of the subjective value is. Thus, those people who do not consider the subjective value to be very important in their decision process, will have a great chance of making rational transformative decisions. However, perhaps Paul is right, that a central tenet or our culture's ordinary way of thinking about transformative decisions is to place sufficiently high emphasis on the subjective value. We will now see that this view is mistaken.

3 Empirical Studies

Paul claims that transformative decisions d_t are not rational. In order for Paul's argument to go through, the subjective value of a transformative experience needs to be the central decision criterion. Furthermore, in the previous section, we developed a model that allows us to calculate the likelihood of an individual making a rational transformative choice, provided we do have the weights of all the decision criteria of that individual. In this part of the paper, we present the results of an empirical investigation on transformative decisions to examine (i) whether the subjective value of a transformative decision is central, and (ii) how likely it is that subjects make rational transformative choices.

3.1 Methods

Paul and most commentators of her work mostly focus on those transformative decisions that are not only epistemically (e.g., spices case) but also personally transformative, i.e. decisions that will change the core preferences of the person making the decision. The two examples that have arguably received the greatest attention in Paul's work are the decisions on whether or not to become a parent and whether or not to become a vampire. We also

added our own case in which a subject has to decide to either leave earth or continue to live on earth but under alien rule. In what follows we present the results of three empirical studies involving those personally transformative choices.

In order to determine the most plausible criteria for all three cases, we ran a pretest in which we randomly assigned 81 participants to either the parent case, the vampire case, or the alien rule case. Subjects were asked to tell us three criteria they would consider most important when making such a decision. The most common answers were then collected and categorized. In the parent case, the following five criteria were most frequently mentioned: (a) discussions with partner, (b) financial costs, (c) collecting information from books and various other sources, (d) subjective value, (e) consistency with other goals in life. Some participants also mentioned that they would contemplate how the decision would change their life. Interestingly, Paul's positive account relates closely to these responses. According to her positive view, we should base our choices in transformative decisions on whether we will be happy to possibly become a person with fundamentally different preferences and undergo changes to our personality. Given the responses and our interest in empirically investigating Paul's suggestion, we also included openness to change as a sixth decision criteria. We would also like to mention that Antti Kauppinen (2015) has suggested that people should base their transformative choices on how much they are consistent with their previous goals. While Kauppinen makes a normative claim on how people should make a decision, the descriptive version of his claim closely matches criteria (e). Paul has repeatedly argued that the subjective value of an outcome of a decision needs to be understood widely as the "value of lived experience"(Paul, 2015b, p. 513). In the General Discussion we will discuss the objection that some of the criteria that we listed as independent from the subjective value actually fall under Paul's conception of the value of lived experience.

For the main study, a total of 181 participants were recruited on Amazon's

Mechanical Turk. 81 participants were assigned to the parent case, 39 (14 female, $M_{age} = 36.87$, STD = 11.43) participants to the vampire case, and 41 (16 female, (16 female, $M_{age} = 36.61$, STD = 12.20)) participants to the alien rule case. In the parent case, 28 participants had to be excluded because they were already parents. However, we did analyze their ratings to investigate the robustness of the importance of the decision criteria. Of the remaining 53 participants, 23 were female, and the mean age was $M_{age} = 31.87$ (STD = 9.75). The vignette in the parent case read as follows:[18]

Parent Case Imagine considering becoming a parent and having to decide whether or not to have a child. How would you arrive at a decision?

The participants were then asked to rate the importance of six decision criteria (random order) on a scale from '0' meaning 'not at all important' to '10' meaning 'extremely important'. The decision criteria were as follows:

- I discuss it with my partner.
- I imagine what it will feel like to have the experiences and emotions when being a parent.
- I consider external factors like financial costs.
- I consider whether becoming a parent is most consistent with the goals I have for my life, like career plans and my vision of family life.
- I consider whether I will be happy to possibly become a person with fundamentally different preferences and undergo changes to my personality.
- I read about the pros and cons of having a baby.

[18]The vignette in the vampire case, for instance, was formulated following Paul's captivating introduction on p. 1 in Paul (2014). The vignette in the alien rule scenario was: Imagine that aliens have taken over the world, and you have to make the choice between either being transported to an unknown planet where you can start a new life, or staying on earth but living under unknown alien rule. Now suppose you have to make the decision immediately. How would you arrive at a decision?

The subjective value was operationalized by using the criterion "I imagine what it will feel like to have the experiences and emotions when being a parent".

3.2 Results

In order to evaluate the differences between the various decision criteria in relation to the rating for the subjective value, we applied a repeated measures ANOVA with *Rating* as the dependent variable and *Criteria* as independent factor with six levels (ask partner, subjective value, costs, consistency, openness, reading). The means and p-values for the criteria in the parent case are listed in Table 1.

Parent				
Decision Criteria				
	Mean	Std Dev	Mean Difference	p-Value Comparison
Ask Partner	8.66	2.52	-0.98	0.164
Subjective Value	7.68	2.26	x	x
Costs	8.26	2.24	-0.58	1.000
Consistency	7.85	2.69	-0.17	1.000
Openness	6.43	3.04	1.25	0.174
Reading	5.45	3.34	2.23	0.001

Table 1: The importance of all six decision criteria in the parent-case averaged over all participants.

For the parent case, the pairwise comparisons show that the ratings for the subjective value (7.68) is only fourth after ask partner (8.66), costs (8.26), and consistency (7.85). Differences between these four criteria were not significant. A slightly different picture was obtained in the vampire scenario. Here, the subjective value was rated second highest (6.79) after openness to change (7.18). In the alien rule case, the subjective value (7.15) was the most important decision criterion but not significantly different compared to all other decision criteria. Thus, in all three studies we conducted, the subjective value, was only one of many decision criteria that was considered to be important.

Given the absolute ratings of each participant, we calculated the relative weights of all six criteria for each individual.[19] We then determined the average weights of all the criteria across all participants. When comparing the importance of these criteria, we found that the subjective value determines the outcome of the decision process only by 17%. Other criteria like financial costs (19%) and what one's partner wants (20%) seem to be on average more important to people. The results are shown in Figure 2 and demonstrate that the relative weight of the subjective value is much smaller than the cut-off value of 0.5 that we determined to be the crucial value for Paul's argument to be sound.

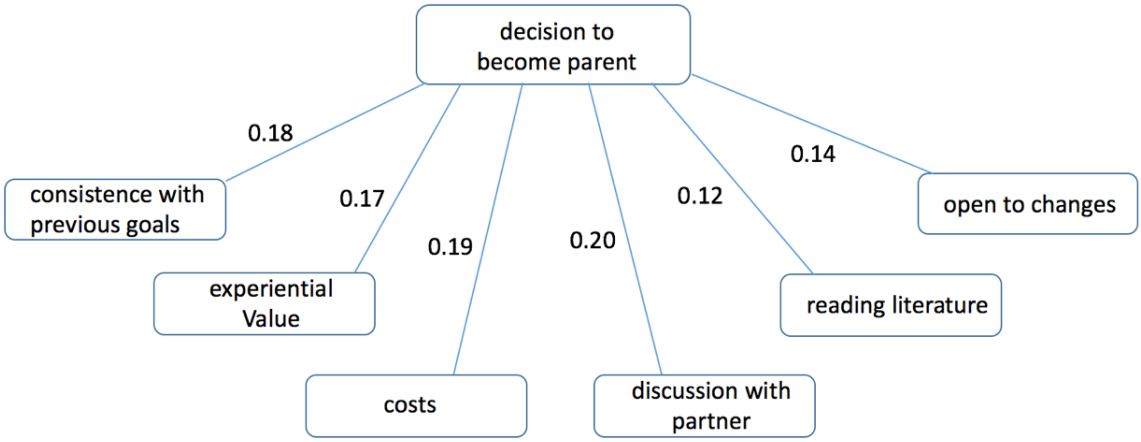

Figure 2: Relative weights of all decision criteria averaged over all 52 participants in the parent-case. The subjective value only takes 17% of the overall weight in the decision on whether or not to become a parent and is not more central than other criteria.

[19]The relative weights of the decision criteria were calculated as follows: We divided each individual rating by the sum of the ratings of all six decision criteria. Thus, imagine that a person gave the following ratings: Ask partner (7), Subjective Value (6), Costs (8), Consistency (5), Openness (3); Reading (1). The sum of all ratings is hence 30. The relative weights of the decision criteria are consequently: Ask partner (7/30 = 0.23), Subjective value (0.2), Costs (0.27), Consistency (0.17), Openness (0.1); Reading (0.03). Note also, that the results do not change when irrelevant criteria are added because the overall sum of the ratings is independent of the number of criteria.

3.3 Likelihood of Rational Decision

Although we have established that the subjective value is not central and hence the decision process possibly rational, we have argued in Section 2, that the weights themselves do not give us a final verdict of whether a subject will make a rational decision or not. To make that determination, we need the respective values of the decision criteria, i.e. we need to know whether a certain criterion speaks in favor or against a certain outcome. As we do not know these values, primarily because they are very difficult to come by, we developed a statistical model in Section 2 that enables us to calculate the likelihood that a subject will make a rational choice. By using a binary system, assigning either the values '1' or '0' to the two possible outcomes, we can specify how likely it is that the person will make a rational choice. In order to do so, we took the connection weights of each participant in our study, and calculated (using a computer script) the ratio of rational and not-rational choices (for details see Section 2). As there are five decision criteria aside from the subjective value, the total possible combinations were 32 for each participant.

Applying the statistical model revealed that only 6 out of 53 participants who completed the survey had a chance of less than 50% to make a rational choice. The distribution of the number of rational decisions for each participant in the study is depicted in Figure 3. On average, the subjective value influences people's decisions in only 35% of cases. Thus, people have a chance of around 65% to make a rational choice. (For comparison: discussions with partner determine almost 50% of all decisions, reading literature in only 24% of the possible combinations). In the alien rule case, in which the subjective received the highest rating, the results do not support Paul's thesis much more strongly: 59% of the possible decisions turned out to be rational decisions. It is therefore safe to say that given the empirical results we obtained and the statistical model we applied, it is likely that people make rational transformative decisions. This is not because we exclude the subjective value

from the decision-making process like the king who needs an heir. Instead, the subjective value is contemplated about but its role can often be neglected due its relatively low significance.

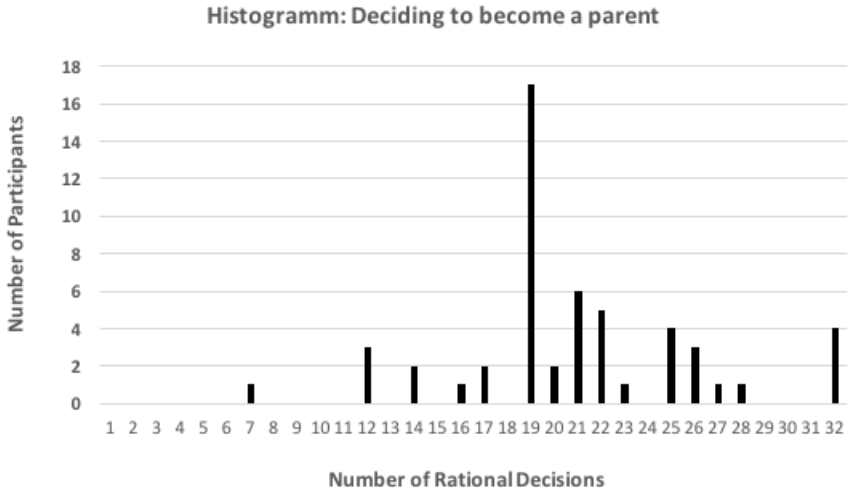

Figure 3: Histogram showing how many of a total of 32 decisions turn out to be rational for each participant using the statistical model described in the main text, e.g. one participant rated the importance of the criteria such that only 7 out of 32 decisions would be rational, for four participants, all combination yielded rational decisions.

3.4 Discussion

The decision to become a parent is one of the most wide-ranging decisions a person can make in his or her life. When we make this decision, we take various decision criteria under consideration, among them the subjective value of what it is like to be a parent. Other transformative decisions like whether or not to become a vampire also feature largely in discussions surrounding Paul's claim on the rationality of transformative decisions. The results of the empirical studies demonstrate that the subjective value fails to play the central role in these decision processes for most people. Instead, as the average values suggest, the subjective value is only one among many decision criteria that plays a certain role when making such a decision. Thus, it seems that

the claim that transformative decisions are not rational needs to be substantially revised. In fact, most people when facing transformative decisions – assuming our studies to be representative – have a very good chance that their decision will be rational.

4 General Discussion

Some of the most fundamental decisions we make in our lives are transformative and – if Paul's argument is sound – fall outside the realm of rationality. Thus, the importance of a project investigating transformative decisions both empirically as well as analytically can hardly be underestimated. In this paper, we have done three things: First, we have highlighted a crucial premise in Paul's argument, the investigation of which has so far been neglected. In order to be able to conclude that transformative decisions are not rational, the subjective value of a transformative decision needs to be *central* for the decision process. Second, we have specified the conditions under which transformative decisions are *possibly* rational and *likely* to be rational. Our analysis showed that if the weight of the subjective value is lower than the weights of all other criteria taken together, the transformative decision is possibly rational. We then developed four statistical models which suggest that the lower the importance of the subjective value, the more likely it is that a transformative decision is rational. Third, the empirical investigation on transformative decisions revealed that the importance of the subjective value is often much lower than has so far been assumed both by Laurie Paul herself as well as other scholars working on this topic. For most people, it is not only the case that it is possible that they make a rational transformative choice but it is also quite likely that they do so.

It is important to note that we have not determined how many people *actually decide* rationally when facing transformative choices. The experiments were not designed to deliver such an answer. In order to make that

determination, we have argued in Section 2 that not only does one need to collect the weights w_i of the various decision criteria (which we have done), but also the values v_i that indicate whether the respective criteria speak in favor or against a certain outcome (which we have not done). Experiments that collect these values v_i are highly desirable but are also much more costly and time-consuming to come by.[20] However, given the statistical models we developed, it is still possible to move beyond what is actually given in the empirical data. For each and every participant in our study, the application of such models allowed us to calculate the likelihood with which they will make a rational choice. The results pose a serious challenge to Paul's depiction of transformative choices. While she would be right to claim that transformative decisions are possibly not rational, the empirical reality reveals a picture according to which transformative decisions are likely to be rational for the majority of people. In the remainder of this paper, we tackle four objections against both our approach as well as the results we presented.

First, one might argue that whereas we have tackled the question of the centrality of the subjective value, we have neglected the role of possible changes in a person's core preferences, that are a center piece in Paul's discussion. However, note, that the issue of possible changes in a person's core preferences only becomes a problem, if the decision is at least partially based on the subjective value itself. Thus, a person who makes a decision in favor of becoming a mother and brackets the transformative experience from her considerations, does not make a non-rational choice even if she later regrets her decision based on the experiences she makes.[21] Paul herself has stated that people who decide to become a parent because, e.g., they want more hands

[20] One possible exception to the claim that we do not know the direction of the criteria are the financial costs of having a child, which in our culture will mostly speak against having a child. The decision-making problem could therefore perhaps be simplified.

[21] See Pettigrew (2015) for an interesting discussion of the problem of how to weigh preferences, which might change as a result of making a choice. In this regard there might also exist an interesting question for personal identity (Parfit 1984), i.e., do transformative experiences undermine psychological continuity?

to work the farm, do not fail to make rational decisions. Nonetheless, one might push this objection a little further by arguing that changes to one's preferences not only affect the values v_i which we assign for each criterion to different outcomes, but also the weights w_i of the various criteria themselves. Furthermore, although many people in our studies did not rate the subjective value to be sufficiently high to have an impact on their decision, the subjective value was not completely disregarded. Thus, if a person's preferences change such that the weight of the what-it-is-like criterion becomes so important that she would have decided differently, then arguably the decision would have failed to be rational. In the end, we believe, this objection is also open to empirical verification. While we do not have hard evidence for how the relative importance of decision criteria in transformative choices might change, the data we collected from people who already have children indicate that the weights of the decision criteria (at least for the parent case) are fairly robust (see also next objection).

A second objection can be mounted against the general approach that we pursue in this paper. How can we be certain that the participants' responses reflect their real preferences or their real behavior? It is known from psychological studies, for instance, that moral responses in controlled settings do not always correlate with real moral behavior (Epley et al. 2000). However, a variety of empirical studies have also shown high consistency between people's stated preferences and their revealed preferences, e.g., when it comes to consumer choices (Loureiro et al. 2003) or family planning (De Silva 1991). Thus, the strength of the objection is likely to strongly depend on the respective context. Unfortunately, there is no empirical evidence we are aware of that points in favor of or against using our methodology for the investigation of transformative choices. We are also not aware of any theoretical arguments from the literature for why people's responses might not reflect their real preferences in our experimental settings. It might, of course, be argued that there is a large time gap between people's revealed preferences

and when they finally make such decisions. A shorter time gap could disclose a greater importance of the subjective value. Let us give a two-pronged response. First, we can safely assume that all the people we asked have already made many transformative decisions – some more important, some less. It is therefore quite plausible to expect that those people can draw upon a wealth of past experiences when judging how important the subjective value is for upcoming decisions. Second, for the parent case, we not only asked participants who are not yet parents, but also those who are already parents. We asked the latter group to tell us how important the various criteria were, when they made their decision. The results reveal a remarkable consistency between non-parents and parents. The only criterion for which a significant difference was recorded was <costs>. Whereas singles rate the importance of financial costs highly, parents do not believe that those costs were crucial in their decision to become parents.

The third objection tackles a crucial underlying presupposition in our approach. A strong assumption of multi-criteria decision processes is that a proper evaluation of the various criteria, requires that they are independent of each other. This independence is important because a criterion should not be *double-counted*. The experiments that we conducted do not discuss whether people actually make trade-offs between different criteria. To illustrate the problem, imagine that two criteria (e.g. costs and consistency with previous goals) speak against having a child. If people interpret consistency with goals as having a specific amount of money, then costs are at least partially double-counted. In other words, the criteria which speak against an outcome would then be overstated. While we have taken care to minimize such distortions by conducting pretests to select the respective criteria, it is hardly possible to fully avoid small dependencies in individual responses.

Fourth, and relatedly, one might challenge the idea that the importance of the what-it-is-like criteria can be determined independently from the other criteria. One might even go so far in arguing that the what-it-is-like criterion

functions as a placeholder for every other criteria, e.g. when you imagine what it will feel like to have a child, you imagine what it will feel like to spend less money for holidays, etc. Paul has explicitly stated that the subjective value includes "ways we'd experience ourselves in such outcomes"(2015b, p. 514). Kauppinen (2015) and Paul (2015b) have, for instance, debated in how far the criterion <consistency with goals> is an integral part of the subjective value in a transformative decision. However, it seems that at least in the parent case the majority of the criteria like <ask partner>, <costs>, and <reading> have little to no bearing on the subjective value of the outcomes. Thus, independently of how you imagine it will feel like, you either consider your partner's opinion and the ensuing costs of the decision to be important or not. We can also assume that the criterion <openness to change> is distinct from the subjective value, given that Paul suggests we should rather make transformative decisions by relying on how much we value new experiences or new selves instead of relying on the subjective value of the outcomes.

Investigating transformative decisions empirically comes with several methodological difficulties. We hope that further experimental studies will follow, which refine the methods and hence deliver more precise results in regards to when and under which conditions people make or are likely to make a rational transformative decision. Given the results we presented in this paper, it seems highly likely that many transformative decisions fall inside the realm of rationality. We believe the least we have done is to shift the burden of proof to those who continue to argue for the opposite claim.

5 Conclusion

The status of transformative experiences plays a crucial role in the argument against the rationality of transformative decisions. Not only are those experiences supposed to be impenetrable, the subjective value needs to be central in transformative decision processes. In this paper, we have analyzed the notion

of centrality and have shown what it takes for transformative decisions to be possibly as well as likely rational. We then provided an empirical investigation of transformative decisions. Perhaps surprisingly, the subjective value not only turned out to be not central in transformative decision processes, the results revealed that most people are more likely than not to make rational transformative decisions even if they take into account the subjective value of the transformative choice.

References

Campbell, J. (2015). L.A. Paul's Transformative Experience, *Philosophy and Phenomenological Research*, 91 (3), 787-793.

Cappelen, H. & Dever, J. (2016). Empathy and transformative experience without the first person point of view, *Inquiry* 60 (3), 315-336.

Chang, R. (2015): Comparativism: The Ground of Rational Choice, in: Weighing Reasons, E. Lord und B. McGuire (Hg.), Oxford: Oxford University Press.

De Silva, I. (1991). Consistency Between Reproductive Preferences and Behavior: The Sri Lankan Experience. *Studies in Family Planning*, 22 (3), 188-197.

Dougherty, T., Horowitz, S., Sliwa, P. (2015). Expecting the Unexpected. *Res Philosophica*, 92 (2), 301-321.

Epley, N., Dunning, D. (2000). Feeling 'Holier Than Thou': Are Self-Serving Assessments Produced by Errors in Self- or Social Prediction? *Journal of Personality and Social Psychology*, 79 (6), 861-875.

Friedman, J. (unpublished manuscribt). Epistemically Transformative Experience.

Ishizaka, A. & Nemery, P. (2013): *Multi-Criteria Decision Analysis*. Singapore: Wiley.

Kauppinen, A. (2015). What's So Great About Experience?. *Res Philosophica*, 92 (2), 371-388.

Loureiro, M. & McCluskey, J. & Mittelhammer, R. (2003). Are Stated Preferences Good Predictors of Market Behavior?. *Land Economics*, 79 (1), 44-55.

Messerli, M. & Reuter, K. (2017). Hard Cases of Comparison. *Philosophical Studies*, 174(9), 2227-2250.

Parfit, D. (1984). *Reasons and Persons*, Oxford: Oxford University Press.

Paul, L.A. (2014). *Transformative Experience*. Oxford: Oxford University Press.

Paul, L.A. (2015a). What You Can't Expect When You'Re Expecting. *Res Philosophica*, 92 (2), 1-23.

Paul, L. A. (2015b). Transformative Choice: Discussion and Replies. *Res Philosophica*, 92 (2), 473-545.

Pettigrew, R. (2015). Transformative Experience and Decision Theory. *Philosophy and Phenomenological Research*, 91 (3), 766-774.

Savage, L. (1954). *The Foundations of Statistics*. New York: John Wiley Sons Inc.

Simon, H. (1953). *A behavioral model of rational choice*. Santa Monica: Rand.

Shupe, E. (2016). Transformative Experience and the Limits of Revelation. Philosophical Studies 173 (11), 3119-3132.

Weber, M. (2000). A New Defense of Satisficing, in: Satisficing and Maximizing, Byron, M. (Hg.), Cambridge: Cambridge University Press.